Employee Attrition Prediction Report

# 1. Objective

To build a classification model that predicts whether an employee will leave the company, and to provide explainable, data-driven recommendations for HR retention strategies.

# 2. Dataset Description & Preprocessing

* Dataset Overview:

- Source: IBM HR Analytics Employee Attrition & Performance

- Records: ~1,470 employees

- Features include: job role, age, gender, income, job satisfaction, overtime, etc.

- Target variable: Attrition (Yes/No)

* Preprocessing Steps:

- Dropped non-informative columns: EmployeeNumber, Over18, StandardHours, EmployeeCount

- Converted Attrition to numeric (Yes → 1, No → 0)

- Applied pd.get\_dummies() to encode categorical variables

- Handled missing values using mode or by dropping rows

- Split the data into train/test sets using stratified sampling

# 3. Models Implemented & Evaluation

* Random Forest Classifier:

- Selected for its robustness and capability to handle both categorical and numerical features

- Used default parameters with random\_state=42

* Evaluation Metrics:

- Accuracy

- Classification Report (Precision, Recall, F1-Score)

- Confusion Matrix

# 4. Key Visualizations & Insights

* Insights from EDA:

- Higher attrition among employees with lower MonthlyIncome

- Employees with Overtime = Yes more likely to leave

- Lower JobSatisfaction and WorkLifeBalance correlate with higher attrition

* SHAP Explanation:

- Used shap.TreeExplainer to explain model predictions

- Top influential features: OverTime, MonthlyIncome, JobSatisfaction, Age, TotalWorkingYears

- SHAP summary plot helped visualize global feature importance

# 5. Actionable Insights for HR

|  |  |
| --- | --- |
| Insight | HR Recommendation |
| Overtime is highly linked to attrition | Review workload policies and reduce overtime |
| Low monthly income increases attrition | Consider compensation restructuring or bonuses |
| Low job satisfaction is a strong signal | Improve team engagement and conduct regular feedback reviews |
| Younger employees tend to leave more | Launch mentorship and growth programs for young talent |
| Poor work-life balance correlates with attrition | Promote flexible work and mental health initiatives |

# 6. Challenges Faced and Solutions

|  |  |  |
| --- | --- | --- |
| Challenge | Description | Solution Implemented |
| Categorical Variables Caused Errors | Error: could not convert string to float | Used pd.get\_dummies() to encode categorical variables |
| Missing Columns During Drop | KeyError for non-existing columns | Verified column names before dropping |
| SHAP Plot Error | SHAP expected matrix but received vector | Removed incorrect indexing and passed full SHAP values |
| IndentationError | Code failed due to bad indentation | Corrected indentation using consistent spacing |
| Target Variable NaNs | NaNs in Attrition column | Filled NaNs using mode or dropped rows |
| Class Imbalance | Attrition = Yes only ~16% | Used stratify=y during train\_test\_split() |
| Explaining Predictions | Hard to interpret results for HR | Used SHAP plots for visual explanations |